2015

Gene Silencing in Microalgae: Mechanisms and Biological Roles

Eun-Jeong Kim
University of Nebraska-Lincoln, sassyej@gmail.com

Xinrong Ma
University of Nebraska-Lincoln

Heriberto Cerutti
University of Nebraska-Lincoln, hcerutti1@unl.edu

Published in *Bioresource Technology* 184 (2015), pp. 23–32; doi: 10.1016/j.biortech.2014.10.119

Submitted September 1, 2014; revised October 22, 2014; accepted October 24, 2014; published online October 30, 2014.

Gene Silencing in Microalgae: Mechanisms and Biological Roles

Eun-Jeong Kim, Xinrong Ma, and Heriberto Cerutti

School of Biological Sciences and Center for Plant Science Innovation, University of Nebraska–Lincoln, Lincoln, Nebraska, USA

Corresponding author – Heriberto Cerutti, School of Biological Sciences and Center for Plant Science Innovation, University of Nebraska–Lincoln, E211 Beadle Center, Lincoln, Nebraska 68588-0666, USA; telephone 402-472-0247, email hcerutti1@unl.edu

Abstract

Microalgae exhibit enormous diversity and can potentially contribute to the production of biofuels and high value compounds. However, for most species, our knowledge of their physiology, metabolism, and gene regulation is fairly limited. In eukaryotes, gene silencing mechanisms play important roles in both the reversible repression of genes that are required only in certain contexts and the suppression of genome invaders such at transposons. The recent sequencing of several algal genomes is providing insights into the complexity of these mechanisms in microalgae. Collectively, glaucophyte, red, and green microalgae contain the machineries involved in repressive histone H3 lysine methylation, DNA cytosine methylation, and RNA interference. However, individual species often only have subsets of these gene-silencing mechanisms. Moreover, current evidence suggests that algal silencing systems function in transposon and transgene repression but their role(s) in gene regulation or other cellular processes remains virtually unexplored, hindering rational genetic engineering efforts.

Keywords: epigenetics, histone post-translational modifications, DNA methylation, RNA interference, miRNAs

1. Introduction

Algae are a diverse group of eukaryotic organisms with important roles in marine, freshwater, and terrestrial ecosystems (Worden and Allen, 2010; Tirichine and Bowler, 2011).

The great potential of algae as feedstocks for renewable biofuel and biomaterial production is also gaining recognition (Hu et al., 2008; Radakovits et al., 2010; Gimpel et al., 2013; Leite et al., 2013). Microalgae are microscopic organisms capable of harnessing sunlight and CO_2 to synthesize useful chemical compounds, such as lipids and carbohydrates, which can be converted into fuels and other bio products. However, production of algae-based fuels is technically, but not yet economically, feasible (Lee, 2011; Chisti, 2013). The major economic bottlenecks cited in the literature include microalgae biological productivity, culture systems, crop protection, and harvesting/extraction processes (Hu et al., 2008; Chisti, 2013; Gimpel et al., 2013; Leite et al., 2013).

For large-scale fuel production reliant on algal photosynthesis key objectives will be achieving high productivity per unit of area, environmental (biotic and abiotic) stress tolerance, ease of harvesting and extraction, and a biomass profile optimized for biofuel conversion (Griffiths and Harrison, 2009; Radakovits et al., 2010). However, identifying in nature microalgal strains simultaneously endowed with all these traits has proven difficult (Hu et al., 2008; Griffiths and Harrison, 2009). Additionally, there has been limited success in increasing biomass productivity or oil content in algae by the genetic engineering of individual genes (Radakovits et al., 2010; La Russa et al., 2012; Gimpel et al., 2013), and this limitation emphasizes the importance of comprehending on a genome scale the metabolic and regulatory networks involved in these processes. Indeed, a significant barrier to advancement is that our knowledge of gene function and regulation is still fairly incomplete in most microalgae (Radakovits et al., 2010; Worden and Allen, 2010; Tirichine and Bowler, 2011). In this context, the study of algal gene silencing mechanisms may provide insights into the control of gene expression as well as facilitate the development of tools for rational genetic engineering.

The regulation of gene expression in eukaryotes involves complex mechanisms, operating at the transcriptional and posttranscriptional levels. Chromatin organization modulates the access of regulatory proteins to DNA and influences multiple aspects of transcription and other DNA-related processes (Bannister and Kouzarides, 2011; Ohsawa et al., 2013). Eukaryotic genomes are commonly organized into several types of chromatin, with euchromatin consisting of transcriptionally permissive or active domains and heterochromatin being characterized by densely packed silent regions (Casas-Mollano et al., 2007; Krauss, 2008; Bannister and Kouzarides, 2011). These functionally and structurally different chromatin states are marked by distinct covalent modifications on the DNA and on specific amino acid residues of the nucleosomal histones (Casas-Mollano et al., 2007; Bannister and Kouzarides, 2011; Saze and Kakutani, 2011). For instance, di- or trimethylation of histone H3 lysine 9 (H3K9) or of histone H3 lysine 27 (H3K27) is often associated with silenced chromatin (Krauss, 2008; Shaver et al., 2010; Bannister and Kouzarides, 2011; Saze and Kakutani, 2011; Derkacheva and Hennig, 2014). DNA cytosine methylation also plays a role in repression and in some organisms there appears to be a complex interplay between histone tail modifications and DNA methylation in establishing a silent chromatin structure (Law and Jacobsen, 2010; Saze and Kakutani, 2011; Du et al., 2012; Zhong et al., 2014).

RNA-directed mechanisms have also been co-opted by evolution to generate a broad spectrum of gene regulatory pathways. RNA-mediated silencing is a conserved process in eukaryotes by which small RNAs (~20–30 nucleotides in length) induce the inactivation of

cognate sequences through a variety of mechanisms, including translation inhibition, RNA degradation, and/or transcriptional repression (Cerutti and Casas-Mollano, 2006; Carthew and Sontheimer, 2009; Meister, 2013). The function of long double-stranded RNAs, as precursors of small RNAs, in triggering gene silencing was initially characterized in *Caenorhabditis elegans* and termed RNA interference (RNAi) (Fire et al., 1998). Yet, in slightly over a decade, RNAi has evolved into a fascinating biological phenomenon intersecting with multiple cellular pathways. Indeed, histone post-translational modifications, DNA cytosine methylation, and RNA-mediated mechanisms impinge on many cellular processes including, besides regulation of gene expression, DNA repair and recombination, chromatin structure, chromosome condensation/stability, as well as the suppression of viruses and transposable elements (Cerutti and Casas-Mollano, 2006; Carthew and Sontheimer, 2009; Cerutti et al., 2011; Ohsawa et al., 2013; Oliver et al., 2014). Moreover, gene silencing mechanisms seem to be important for the integration of environmental and intrinsic stimuli in the control of gene expression and their disruption leads to physiological and developmental abnormalities (Bannister and Kouzarides, 2011; Ohsawa et al., 2013).

In most algal species, both chromatin-associated and RNA-mediated silencing pathways remain largely uncharacterized, even at the level of identifying crucial gene factors in the sequenced genomes. This review will examine the existence of key histone lysine methyltransferases, DNA cytosine methyltransferases, and core components of the RNA-mediated silencing machinery in microalgae. However, algae are very diverse phylogenetically (Worden and Allen, 2010; Tirichine and Bowler, 2011) and, to simplify the identification of commonalities, the analysis will be restricted to microalgae in the Archaeplastida eukaryotic super group, which includes glaucophytes (Glaucophyta), red algae (Rhodophyta), green algae (Chlorophyta), as well as land plants (Streptophyta) (Table 1). We will also discuss briefly the known or inferred biological role(s) of gene-silencing mechanisms in these aquatic organisms. It is anticipated that advances in the basic understanding of gene regulatory mechanisms in microalgae will enable optimization of metabolic pathways of interest through hypothesis-driven genetic engineering strategies.

2. H3K9 and H3K27 methyltransferases in microalgae

2.1. Phylogenetic analysis and domain organization of histone methyltransferases

The methylation of lysine residues in histones, with the exception of H3K79 methylation, is carried out by enzymes that contain an evolutionary conserved SET domain, named after three Drosophila genes (*Su(var)3-9*, *Enhancer of zeste*, and *Trithorax*) (Casas-Mollano et al., 2007; Bannister and Kouzarides, 2011; Huang et al., 2011; Derkacheva and Hennig, 2014). The SET domain constitutes the catalytic site of these lysine methyltransferases (KMTs), but flanking sequences, more distant protein domains, and possibly some cofactors are also important for enzyme activity and specificity (Huang et al., 2011; Krishnan et al., 2011; Derkacheva and Hennig, 2014). To begin characterizing the occurrence and the role(s) of H3K9 and/or H3K27 methyltransferases in microalgae, we surveyed 14 complete or near-complete algal genomes in the Archaeplastida super group for the presence of SETdomain polypeptides (Table 1).

Table 1. Distribution of core gene silencing components in Archaeplastida microalgae

Species	Genome size (Mb)	AGO-Piwi	Dicer	RDR	DNA methyltransferases				Histone H3 methyltransferases		References[b]
					Dnmt1	CMT	Others[a]	Dnmt3	KMT1	KMT6	
Glaucophyta											
Cyanophora paradoxa	70.0	0[c]	0	1	0	0	1	0	0	0	9
Rhodophyta											
Galdieria sulphuraria	13.7	0	0	0	0	0	0	1	0	1	11
Cyanidioschyzon merolae	16.5	0	0	0	0	0	0	1	0	1	5
Porphyridium purpureum	19.7	3	1	3	0	0	0	0	0	0	1
Chlorophyta											
Ostreococcus tauri	12.5	0	0	0	0	0	0	0	0	1	4
Ostreococcus lucimarinus	13.2	0	0	0	0	0	0	0	0	1	8
Bathycoccus prasinos	15.1	0	0	0	0	0	0	0	0	1	7
Micromonas sp. RCC299	20.9	1	0	0	0	0	0	0	0	1	13
Micromonas pusilla CCMP1545	21.9	0	0	0	0	0	0	0	1	1	13
Chlorella variabilis NC64A	46.0	1	1	0	1	1	1	0	2	1	2
Chlorella sorokiniana	56.8	1	1	0	1	1	1	0	2	1	Draft
Coccomyxa subellipsoidea	48.8	2	1?[d]	1	0	0	2	0	2	1	3
Chlamydomonas reinhardtii	120.0	3	3	0	3[e]	0	1	0	1	1	6
Volvox carteri	138.0	2	1	0	1	0	0	0	1	1	10
Streptophyta											
Arabidopsis thaliana	125.0	10	4	6	4	3	0	2	10	3	12

a. DNA methyltransferases that cannot be clearly categorized (see text for details).
b. References: 1, Bhattacharya et al., 2013; 2, Blanc et al., 2010; 3, Blanc et al., 2012; 4, Derelle et al., 2006; 5, Matsuzaki et al., 2004; 6, Merchant et al., 2007; 7, Moreau et al., 2012; 8, Palenik et al., 2007; 9, Price et al., 2012; 10, Prochnik et al., 2010; 11, Schönknecht et al., 2013; 12, Arabidopsis Genome Initiative, 2000; 13, Worden et al., 2009.
c. Total number of genes in the genome encoding a certain factor.
d. ?, predicted protein fairly divergent from canonical model.
e. Includes the chloroplast targeted DMT1 methyltransferase and its putative paralog (see text for details).

Proteins with conserved SET domains were identified by either BLAST or PSI-BLAST searches of protein and/or translated genomic DNA databases, using as queries known *Arabidopsis thaliana* or *Homo sapiens* polypeptides containing SET motifs. Since several of the examined genomes are in draft stage, an important caveat in our analyses is that some proteins may be missing from the databases whereas others may have errors in the predicted gene structure. However, with few exceptions, we considered as potential homologs only proteins that exhibited enough sequence similarity to be aligned and used for phylogenetic tree construction. Interestingly, phylogenetic analysis of the extracted SET-domain proteins revealed that they could be grouped into several distinct classes (see also Huang et al., 2011), but we examined in detail only KMT1 and KMT6 homologs (Figs. 1 and 2), following the nomenclature proposed for yeast and metazoan lysine methyltransferases (Allis et al., 2007).

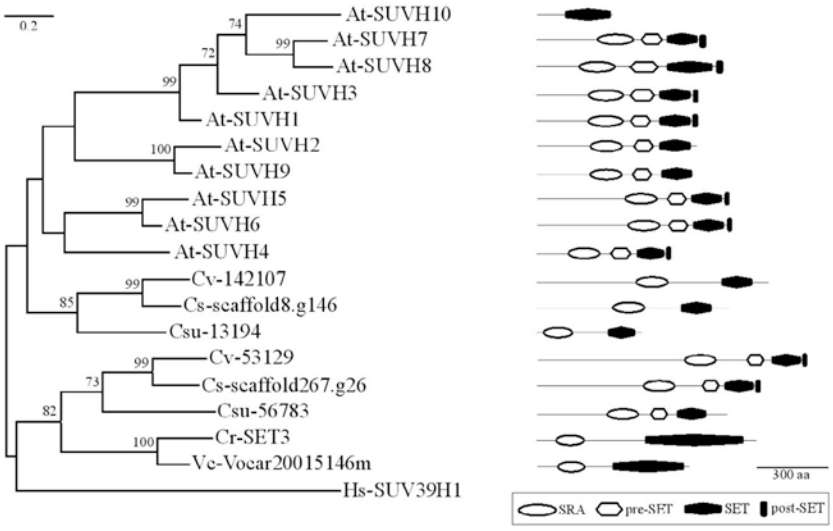

Figure 1. Maximum likelihood tree showing the phylogenetic relationship among KMT1 proteins. Sequences corresponding to KMT1 proteins from different organisms were aligned using Muscle, and the tree was drawn using the MEGA v6.06 program. Numbers indicate bootstrap values, > 60%, based on 1000 pseudoreplicates. Species are designated by a two- or three-letter abbreviation preceding the name of each protein. At, *Arabidopsis thaliana*; Cr, *Chlamydomonas reinhardtii*; Cs, *Chlorella sorokiniana*; Csu, *Coccomyxa subellipsoidea* C-169; Cv, *Chlorella variabilis* NC64A; Hs, *Homo sapiens*; Vc, *Volvox carteri*. Accession numbers of proteins used to draw the tree are: At-SUVH1 to At-SUVH9, AAK28966 to AAK28974, respectively; At-SUVH10, NP_178647; Cr-SET3, XP_001701764; Cs-scaffold8.g146 and Cs-scaffold267.g26, from *Chlorella sorokiniana* draft genomic sequence assembly at the University of Nebraska–Lincoln; Csu-13194, XP_005649968; Csu-56783, XP_005646116; Cv-142107, XP_005850072; Cv-53129, XP_005846419; Hs-SUV39H1, NP_001269095; Vc-Vocar20015146m, XP_002952737. The SMART 7 database was used to identify conserved protein motifs and the domain organization of the proteins is indicated on the right (see text for details).

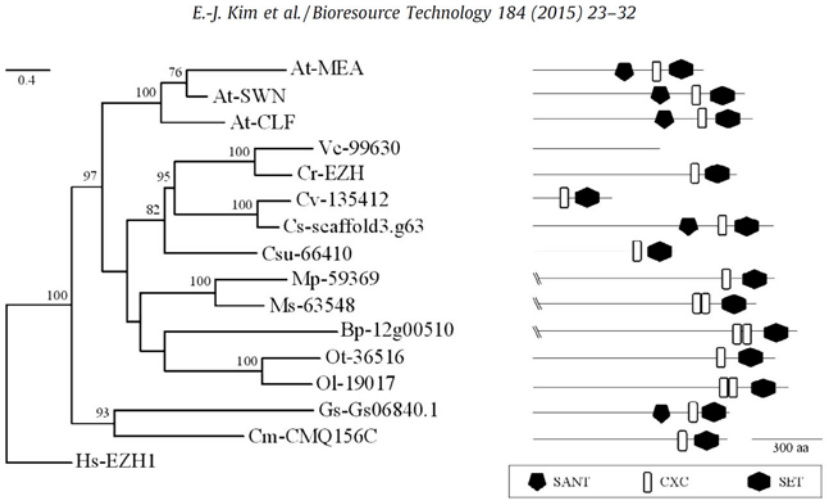

Figure 2. Maximum likelihood tree showing the phylogenetic relationship among KMT6 proteins. Sequences corresponding to KMT6 proteins from different organisms were aligned using Muscle and the tree was drawn using the MEGA v6.06 program. Numbers indicate bootstrap values, > 60%, based on 1000 pseudoreplicates. Species are designated by a two-letter or three-letter abbreviation preceding the name of each protein, as described in the legend to Figure 1, except for: Bp, *Bathycoccus prasinos*; Cm, *Cyanidioschyzon merolae*; Gs, *Galdieria sulphuraria*; Mp, *Micromonas pusilla* CCMP1545; Ms, *Micromonas* sp. RCC299; Ol, *Ostreococcus lucimarinus*; Ot, *Ostreococcus tauri*. Accession numbers of proteins used to draw the tree are: At-CLF, NP_179919; At-MEA, NP_563658; At-SWN, NP_567221; Bp-12g00510, XP_007509828; Cm-CMQ156C, XP_005538096; Cr-EZH, Cre17.g746247 at http://phytozome.jgi.doe.gov; Cs-scaffold3.g63, from *Chlorella sorokiniana* draft genomic sequence assembly at the University of Nebraska–Lincoln; Csu-66410, PID 66410 at http://genome.jgi.doe.gov/pages/search-for-genes.jsf?organism=Coc_C169_1; Cv-135412, PID 135412 at http://genome.jgipsf.org/ChlNC64A_1/ChlNC64A_1.home.html; Gs-Gs06840.1, XP_005708796; Hs-EZH1, NP_001982; Mp-59369, XP_003059747; Ms-63548, XP_002505475; Ol-19017, XP_001422533; Ot-36516, XP_003082958; Vc-99630, XP_002958369. The SMART 7 database was used to identify conserved protein motifs and the domain organization of the proteins is indicated on the right (see text for details).

Members of the algal KMT1 class, like animal and plant KMT1 proteins, are likely responsible for H3K9 methylation, an epigenetic mark involved in gene silencing and heterochromatin formation (Casas-Mollano et al., 2007; Krauss, 2008; Bannister and Kouzarides, 2011; Huang et al., 2011). In the examined microalgae, KMT1 homologs appear to be limited to species of the Chlorophyta clade, including organisms in the Trebouxiophyceae (*Chlorella sorokiniana*, *Chlorella variabilis* NC64A, and *Coccomyxa subellipsoidea*) and Chlorophyceae (*Chlamydomonas reinhardtii* and *Volvox carteri*) classes (Table 1 and Fig. 1). *Micromonas pusilla* CCMP1545 also seems to code for a KMT1-related protein (Table 1). Yet, the corresponding gene is located in an island of the genome with no detectable homology to the closely related *Micromonas* sp. RCC299 (data not shown) and the functional significance of the encoded protein is currently unclear. Most algal KMT1 proteins show high sequence

similarity to land plant KMT1 polypeptides, both within the SET domain and in the surrounding regions known as the Pre-SET and Post-SET motifs (Fig. 1). Additionally, all algal sequences contain an SRA (SET and RING associated) domain (Fig. 1), which recognizes the methylation status of CG and CHH DNA sequences (where H = A, T, or C) (Rajakumara et al., 2011). Land plant KMT1 proteins have been reported to fall into several distinct subgroups, indicative of functional diversification (Casas-Mollano et al., 2007; Huang et al., 2011). However, our phylogenetic analysis suggests that gene duplication and potential pathway diversification occurred after the divergence of land plants from the lineage leading to green algae (Fig. 1). A phylogenetic tree of KMT6 histone methyltransferases supports a similar conclusion (Fig. 2).

Members of the algal KMT6 class, like animal and plant KMT6 proteins, are likely responsible for H3K27 methylation and putative components of the evolutionarily conserved Polycomb Repressive Complex 2 (Shaver et al., 2010; Bannister and Kouzarides, 2011; Huang et al., 2011; Derkacheva and Hennig, 2014). Within the examined microalgae, KMT6 homologs appear to be widely distributed (Table 1 and Fig. 2). Only the Glaucophyta clade seems to lack this type of histone methyltransferases, with the caveat that the draft genome of *Cyanophora paradoxa* may be incomplete, and this observation may have to be revisited upon genome completion and/or the sequencing of additional species in the clade. Most algal KMT6 polypeptides show high sequence similarity to land plant KMT6 proteins, particularly within the SET domain (Fig. 2). Additionally, almost all algal sequences contain a specific CXC motif (Fig. 2), an ~65-residue cysteine-rich region that has been implicated in DNA binding and shows structural homology to Pre-SET domains (Zheng et al., 2012). However, with two exceptions, algal KMT6 proteins lack the SANT (SWI3, ADA2, N-CoR, and TFIIIB DNA-binding) motif that is found in most land plant homologs (Huang et al., 2011). Overall our findings are consistent, as previously suggested (Shaver et al., 2010), with widespread distribution of KMT6 methyltransferases within algal lineages.

2.2. Biological role(s) of H3K9 and H3K27 methylation

By analogy to their function in higher eukaryotes (Casas-Mollano et al., 2007; Krauss, 2008; Shaver et al., 2010; Bannister and Kouzarides, 2011; Huang et al., 2011; Derkacheva and Hennig, 2014), these post-translational histone modifications are likely involved in transcriptional repression in microalgae. Yet, experimental evidence is almost entirely lacking. To our knowledge, direct determination of H3K9 and H3K27 methylation has only been carried out in *C. reinhardtii*. This alga contains mono- and trimethylated H3K9 as well as mono- and dimethylated H3K27 (Casas-Mollano et al., 2007; Shaver et al., 2010). Additionally, high levels of H3K9 monomethylation have been found associated with transcriptionally repressed transgenes (Casas-Mollano et al., 2007; Strenkert et al., 2013). Interestingly, RNAi-mediated suppression of Chlamydomonas SET3, a KMT1 methyltransferase (Fig. 1), reactivated the expression of silenced repetitive transgenic arrays (Casas-Mollano et al., 2007).

RNAi-mediated suppression of Chlamydomonas EZH, a KMT6 methyltransferase (Fig. 2), also led to defects in the silencing of transgenes and retrotransposons as well as to a global increase in histone post-translational modifications associated with transcriptional activity, such as trimethylation of histone H3 lysine 4 and acetylation of histone H4 (Shaver

et al., 2010). Intriguingly, a *C. variabilis* NC64A virus, *Paramecium bursaria* chlorella virus 1 (PBCV-1), has been shown to encode a KMT6 protein, termed vSET, that might be linked to the rapid inhibition of host transcription after viral infection (Wei and Zhou, 2010). In a heterologous system, vSET causes H3K27 methylation and transcriptional repression (Mujtaba et al., 2008; Wei and Zhou, 2010). These observations, taken together, are consistent with a role of H3K9 and H3K27 methylation in gene silencing in microalgae, possibly as a means to suppress intragenomic parasites such as transposable elements, as well as in host-virus interactions. However, H3K27 methyltransferases appear to have a much broader taxonomic distribution than H3K9 methyltransferases (Table 1), perhaps implicating proteins of the KMT6 class in a more essential, evolutionarily conserved regulatory role(s). Nevertheless, assessing the potential involvement of posttranslational histone modifications in specific gene or pathway regulation in microalgae, as demonstrated in land plant and animals (Bannister and Kouzarides, 2011; Derkacheva and Hennig, 2014), will require further studies.

3. DNA cytosine methyltransferases in microalgae

3.1. Phylogenetic analysis and domain organization of DNA methyltransferases

The DNA of a vast array of organisms contains the modified base 5-methylcytosine, but genomes are methylated in different ways and with different consequences in diverse species (Feng et al., 2010; Zemach et al., 2010; Law and Jacobsen, 2010; Huff and Zilberman, 2014). Introduction of a methyl group at the C5 position of cytosine is catalyzed by a large family of DNA methyltransferases, including six subfamilies characterized by catalytic domains associated with N-terminal or C-terminal extensions containing distinct motifs (Goll and Bestor, 2005; Ponger and Li, 2005; Huff and Zilberman, 2014). In land plants, DNA cytosine methylation can occur in three sequence contexts namely CG, CHG, and CHH (where H = A, T, or C) and three DNA methyltransferase subfamilies have been implicated in the establishment and/or maintenance of methylation at these sequences (Goll and Bestor, 2005; Law and Jacobsen, 2010). CG methylation is maintained by the Dnmt1/MET1 subfamily of DNA methyltransferases, CHG methylation is mediated by the plant-specific chromomethylases (CMTs), and CHH methylation is introduced by the Dnmt3/DRM (Domains Rearranged Methyltransferase) enzymes (Goll and Bestor, 2005; Law and Jacobsen, 2010; Zhong et al., 2014). Interestingly, in a clear interplay between silencing mechanisms, chromomethylases appear to be targeted to chromosomal regions with H3K9 methylated nucleosomes via their bromo adjacent homology (BAH) and chromo domains (Du et al., 2012), whereas DRMs seem to be guided to target loci by small RNAs and certain components of the RNAi machinery (Law and Jacobsen, 2010; Zhong et al., 2014).

The presence of Dnmt1/MET1, CMT, and Dnmt3/DRM polypeptides was examined in the 14 microalgae with sequenced genomes belonging to the Archaeplastida super group, using as queries known *A. thaliana* or *H. sapiens* polypeptides (Table 1). Homologs of Dnmt3/DRM enzymes, which have been implicated in de novo DNA methylation (Goll and Bestor, 2005; Ponger and Li, 2005; Law and Jacobsen, 2010), were found only in the red algae *Galdieria sulphuraria* and *Cyanidioschyzon merolae* (Table 1) and with a structural organization similar to the vertebrate enzymes (data not shown), whereas chromomethylase-

related methyltransferases appear to be restricted to the genus *Chlorella* (Table 1). Dnmt1/MET1 homologs were identified in species of the Trebouxiophyceae (*C. sorokiniana* and *C. variabilis*) and Chlorophyceae (*C. reinhardtii* and *V. carteri*) classes (Table 1 and Fig. 3). The algal Dnmt1/MET1 proteins show high sequence similarity to the land plant polypeptides in the DNA methyltransferase catalytic domain. Additionally, all these algal sequences contain N-terminal extensions with conserved DNMT-RFD (DNA methyltransferase replication foci domain) and BAH domains (Fig. 3), as observed in the canonical enzyme (Goll and Bestor, 2005; Ponger and Li, 2005).

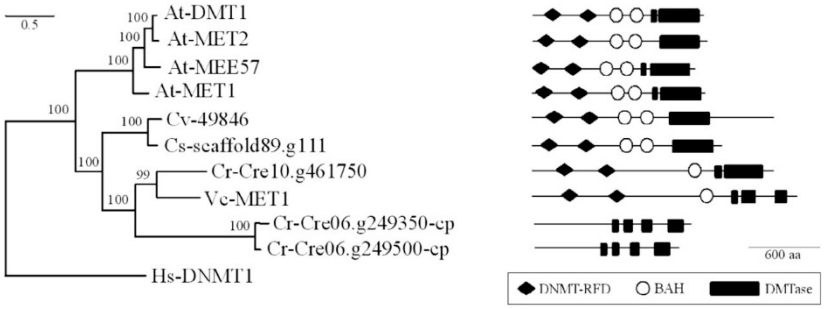

Figure 3. Maximum likelihood tree showing the phylogenetic relationship among Dnmt1/MET1 proteins. Sequences corresponding to the DMTase domain from different organisms were aligned using Muscle, and the tree was drawn using the MEGA v6.06 program. Numbers indicate bootstrap values, > 60%, based on 1000 pseudoreplicates. Species are designated by a two-letter or three-letter abbreviation preceding the name of each protein, as described in the legend to Figure 1. Accession numbers of proteins used to draw the tree are: At-MET1, NP_199727; At-DMT1, NP_192638; At-MET2, NP_001190725; At-MEE57, NP_193097; Cr-Cre10.g461750, Cr-Cre06.g249350-cp, and Cr-Cre06.g249500-cp, Cre10.g461750, Cre06.g249350, and Cre06.g249500 at http://phytozome.jgi.doe.gov; Cs-scaffold89.g111, from *Chlorella sorokiniana* draft genomic sequence assembly at the University of Nebraska–Lincoln; Cv-49846, PID 56675 at http://genome.jgi-psf.org/ChlNC64A_1/ChlNC64A_1.home.html; Hs-DNMT1, NP_001370; Vc-MET1, Vocar20014971m at http://phytozome.jgi.doe.gov. The SMART 7 database was used to identify conserved protein motifs and the domain organization of the proteins is indicated on the right (see text for details).

Interestingly, *C. reinhardtii* contains two additional Dnmt1/MET1 related polypeptides, with high degree of homology to the catalytic domain but lacking conserved motifs in the N-terminal regions (Fig. 3). One of these proteins, termed DMT1, has been characterized as a novel DNA methyltransferase with de novo nonselective cytosine methylation activity (Nishiyama et al., 2004). Moreover, it has been shown to localize to Chlamydomonas chloroplasts and to influence plastid DNA methylation and the uniparental inheritance of chloroplast genes (Nishiyama et al., 2004). Unexpectedly, in several microalgae, we also found putative DNA methyltransferases that cannot be clearly categorized (Table 1, Others). These predicted proteins contain DNA methyltransferase catalytic domains somewhat related to those of the Dnmt1/MET1 or the CMT subfamilies, but they lack either N-terminal

extensions or conserved domains in the N-terminal extensions. If correctly predicted and functional, some of these enzymes might be responsible for DNA methylation patterns or processes unique to microalgae (see below).

3.2. Biological role(s) of DNA cytosine methylation

Methylation of the fifth carbon of cytosine is a widespread modification present in bacteria, archaea, and eukaryotes (Feng et al., 2010; Zemach et al., 2010; Huff and Zilberman, 2014). However, methylation of cytosine residues has not been detected in several model organisms such as *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *C. elegans* (Capuano et al., 2014), suggesting that this modification is not essential for eukaryotic life. Interestingly, some of the examined microalgae lack clearly identifiable homologs of Dnmt1/MET1, CMT, and Dnmt3/DRM methyltransferases (Table 1), but it is not certain that they are devoid of DNA cytosine methylation. For instance, the Prasinophyte algae (*Ostreococcus lucimarinus*, *Bathycoccus prasinos*, and *M. pusilla* CCMP1545) contain highly diverged DNA methyltransferases of the Dnmt5 subfamily and exhibit genomic CG methylation (Huff and Zilberman, 2014). In land plants or mammals, DNA cytosine methylation is often associated with the transcriptional silencing of transposable elements, repetitive DNA sequences, and some protein-coding genes, and plays a critical role in genomic imprinting, X-chromosome inactivation, and chromosome stability (Feng et al., 2010; Law and Jacobsen, 2010). Gene body methylation is also highly conserved even though its precise function(s) remains an open question (Feng et al., 2010; Zemach et al., 2010; Huff and Zilberman, 2014). In contrast to this wealth of information, the role(s) of DNA cytosine methylation in microalgae is still poorly understood.

Advances in sequencing technologies have enabled profiling of the genome methylation patterns of multiple species, including the microalgae *C. variabilis* NC64A, *V. carteri*, and *C. reinhardtii* (Feng et al., 2010; Zemach et al., 2010). In *C. variabilis*, genes are universally CG methylated, mostly within their bodies with a sharp drop at the promoters (Zemach et al., 2010). Interestingly, promoter methylation is inversely correlated with gene expression suggesting that, when it occurs, promoter-proximal methylation represses transcription. CHG methylation is also substantial but, similarly to land plants, concentrated in repetitive (presumably transposon) sequences and excluded from genes (Zemach et al., 2010). In contrast, the *V. carteri* genome is methylated to a much lower degree and exclusively in the CG context. Transposons and repeats are preferentially methylated but a weak negative relationship between promoter methylation and gene transcription was also observed (Zemach et al., 2010). In Volvox, DNA methylation had previously been implicated in the transcriptional silencing of introduced transgenes and a MET1-like methyltransferase was suggested to be involved in the maintenance of transgene and transposon methylation (Babinger et al., 2007).

In *C. reinhardtii*, DNA cytosine methylation was initially observed in chloroplast DNA during gametogenesis and implicated in the uniparental inheritance of mating type plus chloroplast DNA (Umen and Goodenough, 2001; Nishiyama et al., 2004). Recent findings indicated that the nuclear genome of this alga is also methylated at low levels and, similarly to Volvox, CG methylation is preferentially enriched in transposon sequences and to

a much lower degree in gene bodies (Feng et al., 2010). Intriguingly, CHG and CHH methylation was also observed uniformly along chromosomes and showed little enrichment in transposons or repeats (Feng et al., 2010). However, Chlamydomonas does not appear to contain CMT or Dnmt3/DRM homologs (Table 1 and Feng et al., 2010) and the methyltransferase(s) responsible for the later modifications has not been characterized. DNA methylation has also been implicated in the transcriptional silencing of introduced transgenes, particularly tandem repeats, in Chlamydomonas (Cerutti et al., 1997).

Collectively, the evidence suggests that, as in land plants and vertebrates, preferential DNA cytosine methylation of transposable elements and gene bodies has been conserved in some microalgae (Feng et al., 2010; Zemach et al., 2010). The methylation of transposons likely suppresses transcription and ensuing self-replication preserving genome integrity. Additionally, DNA methylation in promoters correlates negatively with gene expression in a subset of microalgae (Zemach et al., 2010), but it is not known whether certain protein genes are regulated by this modification. The pattern of genomic DNA methylation and the complement of DNA methyltransferases in species of the genus Chlorella, except for the lack of a Dnmt3/DRM homolog (Table 1), are similar to those in land plants. In contrast, *C. reinhardtii* and *V. carteri* methylate preferentially transposable elements but predominantly in the CG (rather than the CHG/CHH) context and presumably by a somewhat different mechanism from that in land plants given the divergence in DNA methyltransferases (Table 1). Additionally, *O. lucimarinus*, *M. pusilla*, and *B. prasinos*, which possess DNA methyltransferases of the diverged Dnmt5 subfamily, exhibit densely clustered CG methylation in nucleosome linkers, possibly contributing to nucleosome positioning and proper chromatin compaction in very small nuclei (Huff and Zilberman, 2014). Thus, the distribution and function(s) of DNA cytosine methylation in microalgae appear to be highly varied, making it difficult to extrapolate knowledge from well-characterized model systems as to its possible role(s) in the regulation of gene expression.

4. The RNA interference machinery in microalgae

4.1. Phylogenetic analysis and domain organization of core components of the RNAi machinery

Biochemical and genetic studies in multiple eukaryotes resulted in the identification of three key components of the RNAi machinery, namely Dicer, Argonaute-Piwi (AGO-Piwi), and RNA-dependent RNA Polymerase (RDR) (Cerutti and Casas-Mollano, 2006; Carthew and Sontheimer, 2009; Meister, 2013; Burroughs et al., 2014). In the best-characterized RNAi pathway, a variety of double-stranded RNA (dsRNA) precursors are processed into small RNAs (sRNAs) by the RNaseIII-like endonuclease Dicer (Cerutti and Casas-Mollano, 2006; Carthew and Sontheimer, 2009; Rogers and Chen, 2013). These sRNAs are then incorporated into effector complexes, which include members of the AGO-Piwi family of proteins named after *A. thaliana* ARGONAUTE1 and *Drosophila melanogaster* Piwi (Cerutti and Casas-Mollano, 2006; Meister, 2013; Burroughs et al., 2014). AGO-Piwi proteins act as highly specialized small-RNA-binding modules. Based on structural studies, they are characterized by amino-terminal (N-terminal), PAZ (Piwi, Argonaute, Zwille), MID (middle) and PIWI domains, the latter being related to RNase H (Carthew and Sontheimer, 2009; Meister, 2013; Burroughs et al., 2014). Some AGO-Piwi polypeptides function as sRNA-

guided endonucleases that cleave complementary transcripts, whereas others lack endonucleolytic activity and may be part of complexes involved in nondegradative RNAi mechanisms such as translation repression (Cerutti and Casas-Mollano, 2006; Carthew and Sontheimer, 2009; Meister, 2013; Rogers and Chen, 2013). In certain species, RDRs also play an important role in RNAi, either by producing an initiating dsRNA trigger from single-stranded transcripts or by enhancing the RNAi response through amplification of the sRNA amount (Cerutti and Casas-Mollano, 2006; Burroughs et al., 2014).

The presence of Dicer, AGO-Piwi, and RDR polypeptides was examined in the 14 microalgae with sequenced genomes belonging to the Archaeplastida super group (Table 1). As previously reported (Cerutti and Casas-Mollano, 2006; Cerutti et al., 2011), core RNAi machinery components seem to be entirely absent from several algae with small nuclear genomes such as the red algae *C. merolae* and *G. sulphuraria* and the green algae *O. lucimarinus*, *Ostreococcus tauri*, *B. prasinos*, and *M. pusilla* CCMP1545 (Table 1). *Micromonas* sp. RCC299 appears to code for an AGO-Piwi related protein (Table 1). Yet, the corresponding gene is located in a region of the genome with no detectable homology to the closely related *M. pusilla* CCMP1545 (data not shown), and it is not clear if the encoded protein is actually functional. These observations are consistent with the hypothesis that the RNAi machinery appeared early during eukaryotic evolution and it has been lost independently in multiple lineages (Cerutti and Casas-Mollano, 2006; Cerutti et al., 2011). As a consequence, RNAi machinery components are present in algal species of each of the Glaucophyta, Rhodophyta, and Chlorophyta clades but with patchy distribution (Table 1).

AGO-Piwi proteins can often be identified in microalgae that also contain Dicer-like proteins (Table 1) and the domain organization of AGO-Piwi polypeptides has been well conserved in these species (Fig. 4). In contrast, the multidomain structure of the Dicer enzymes of higher eukaryotes is not well maintained among protists (Cerutti and Casas-Mollano, 2006) and an RNaseIII motif might be all that is strictly required for function (Cerutti and Casas-Mollano, 2006; Patrick et al., 2009). RDR-related proteins are structurally conserved but, consistent with an ancillary role in typical RNAi (Cerutti and Casas-Mollano, 2006), they appear to have a fairly limited taxonomic distribution among the examined microalgae (Table 1). Intriguingly, the glaucophyte *C. paradoxa* only seems to contain an RDR protein, but its incomplete genome prevents a definitive assessment of the RNAi machinery components. Conversely, the red alga *Porphyridium purpureum* and, possibly, the green alga *C. subellipsoidea* encode all three core components of the RNAi machinery (Table 1), although the putative Coccomyxa Dicer is highly diverged (data not shown). In addition, AGO-Piwi encoding genes have undergone duplication in several microalgae (Fig. 4), perhaps associated with diversification of functions as reported in land plants and metazoans (Casas-Mollano et al., 2008; Ghildiyal and Zamore, 2009; Meister, 2013).

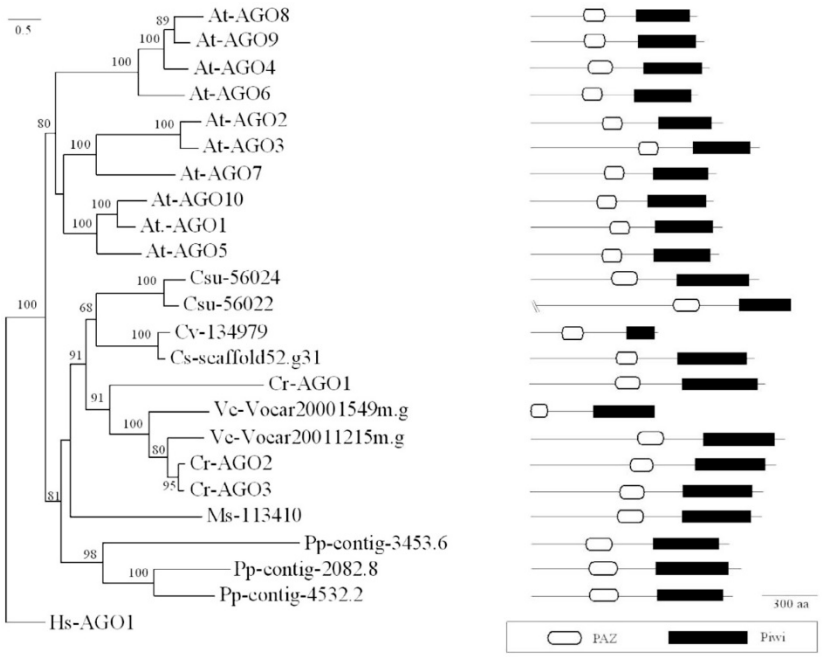

Figure 4. Maximum likelihood tree showing the phylogenetic relationship among Argonaute proteins. Sequences corresponding to AGO-Piwi proteins from different organisms were aligned using Muscle and the tree was drawn using the MEGA v6.06 program. Numbers indicate bootstrap values, > 60%, based on 1000 pseudoreplicates. Species are designated by a two-letter or three-letter abbreviation preceding the name of each protein, as described in the legends to Figures 1 and 2, except for: Pp, *Porphyridium purpureum*. Accession numbers of proteins used to draw the tree are: At-AGO1, NP_849784; At-AGO2, NP_174413; At-AGO3, NP_174414; At-AGO4, NP_001189613; At-AGO5, NP_850110; At-AGO6, NP_180853; At-AGO7, NP_177103; At-AGO8, NP_197602; At-AGO9, NP_197613; At-AGO10, NP_199194; Cr-AGO1, Cr-AGO2, and Cr-AGO3, Cre02.g141050, Cre04.g214250, and Cre04.g689647 at http://phytozome.jgi.doe.gov; Cs-scaffold52.g31, from *Chlorella sorokiniana* draft genomic sequence assembly at the University of Nebraska–Lincoln; Csu-56022 and Csu-56024, PID 56022 and PID 56024 at http://phytozome.jgi.doe.gov; Cv-134979, PID 134979 at http://genome.jgi-psf.org/ChlNC64A_1/ChlNC64A_1.home.html; Hs-AGO1, NP_03633; Ms-113410, PID 113410 at http://phytozome.jgi.doe.gov; Vc-Vocar 20011215m.g and Vc-Vocar20001549m.g, Vocar20011215m and Vocar20001549 at http://phytozome.jgi.doe.gov; Pp-contig-4532.2, Pp-contig-2082.8, and Pp-contig-3453.6, evm.model .contig_4532.2, evm.model.contig_2082.8, and evm.model.contig_3453.6 at http://cyanophora.rutgers.edu. The SMART 7 database was used to identify conserved protein motifs and the domain organization of the proteins is indicated on the right (see text for details).

In many eukaryotes, at least two major classes of small RNAs have been recognized: microRNAs (miRNAs) and small interfering RNAs (siRNAs) (Bartel, 2009; Ghildiyal and Zamore, 2009; Axtell, 2013; Rogers and Chen, 2013). MicroRNAs often modulate gene expression and originate from endogenous, single-stranded noncoding RNA transcripts or introns that fold into imperfect hairpin structures (Bartel, 2009; Ghildiyal and Zamore,

2009; Axtell, 2013; Rogers and Chen, 2013). siRNAs are produced from long, near-perfect complementarity dsRNAs of diverse origins (including dsRNAs experimentally introduced into cells) and play various roles in post-transcriptional regulation of gene expression, suppression of viruses and transposable elements, DNA methylation, and/or heterochromatin formation (Cerutti and Casas-Mollano, 2006; Carthew and Sontheimer, 2009; Ghildiyal and Zamore, 2009; Axtell, 2013; Rogers and Chen, 2013). The occurrence of small RNAs, potentially generated by Dicer-mediated processing, has been examined in very few of the sampled algal species. *C. reinhardtii* has been shown to contain a complex set of endogenous sRNAs including miRNAs, phased siRNAs, as well as siRNAs originating from transposons and repeated DNA regions (Molnar et al., 2007; Zhao et al., 2007; Yamasaki et al., 2013; Voshall et al., 2014). Likewise, *V. carteri* has been demonstrated to possess miRNAs, many of them being preferentially enriched in either gonidia or somatic cells (Li et al., 2014). Interestingly, there appears to be little conservation of miRNA genes among algal species or with those encoded in the genomes of land plants or animals (Cerutti et al., 2011; Li et al., 2014; Voshall et al., 2014).

4.2. Biological role(s) of RNA interference

The presence of core components of the RNAi machinery and of endogenous miRNAs/siRNAs in certain microalgae suggest that RNA-mediated silencing may play a biological role(s) in these organisms. However, it is also apparent that the RNAi machinery seems to be dispensable for cellular life since it has been entirely lost or extensively modified in several species. In a wide range of eukaryotes, sRNA-mediated mechanisms have been implicated in multiple processes including transposon silencing, resistance to viruses, regulation of endogenous gene expression, heterochromatin formation, DNA methylation, DNA repair, and maintenance of genome stability (Cerutti and Casas-Mollano, 2006; Carthew and Sontheimer, 2009; Ohsawa et al., 2013; Rogers and Chen, 2013; Oliver et al., 2014; Zhong et al., 2014). In contrast, very little is known about the actual function(s) of RNA-mediated silencing in microalgae.

C. reinhardtii has undergone duplication of key RNAi components (Casas-Mollano et al., 2008) and contains three Dicers and three Argonautes (Table 1). One of the Dicer proteins, DCL1, appears to be part of a siRNA pathway that has specialized for the control of transposable elements (Casas-Mollano et al., 2008). However, *C. reinhardtii* also relies on a DCL1-independent, transcriptional silencing mechanism(s) for transposon repression (Casas-Mollano et al., 2007; Shaver et al., 2010). Interestingly, this chromatin-mediated silencing is sensitive to temperature, being much more effective at 17°C than at 25°C (Cerutti et al., 1997). Conversely, RNAi dependent post-transcriptional gene repression has been shown to be more efficient at 25–29°C in both invertebrates and land plants (Fortier and Belote, 2000; Szittya et al., 2003). Thus, it is tempting to speculate that in *C. reinhardtii*, multiple, partly independent silencing mechanisms may operate to suppress reliably transposon mobilization over a wide range of environmental conditions (Casas-Mollano et al., 2008; Cerutti et al., 2011). In *C. variabilis* NC64A recent transcriptome analyses revealed that components of certain RNA-silencing pathways are up regulated after infection with the PBCV-1 virus (Rowe et al., 2013). Thus, the limited available evidence does suggest that

RNAi is involved in defense responses against transposable elements and, possibly, viruses in some microalgae.

In higher eukaryotes, RNAi also plays an important role in the regulation of endogenous gene expression through miRNAs and other small RNAs (Bartel, 2009; Ghildiyal and Zamore, 2009; Axtell, 2013; Rogers and Chen, 2013). Within the examined microalgae, candidate miRNAs have been experimentally identified in *C. reinhardtii* (Molnar et al., 2007; Zhao et al., 2007; Yamasaki et al., 2013; Voshall et al., 2014) and in *V. carteri* (Li et al., 2014), and potential endogenous targets have been predicted by computational approaches (Molnar et al., 2007; Zhao et al., 2007; Li et al., 2014; Voshall et al., 2014). However, target gene identification is a challenging problem (Thomas et al., 2010; Voshall et al., 2014) and the false-positive prediction rate for miRNA targets in microalgae is currently unknown. Interestingly, most characterized land plant miRNAs appear to regulate transcripts with highly complementary binding sites and preferentially trigger their endonucleolytic cleavage by Argonaute proteins (Bartel, 2009; Axtell, 2013; Rogers and Chen, 2013). This also appears to be the case for some miRNAs in both *C. reinhardtii* and *V. carteri*, where expected target RNA cleavage products have been experimentally detected (Molnar et al., 2007; Zhao et al., 2007; Li et al., 2014; Voshall et al., 2014). Yet, very few of the reported Chlamydomonas or Volvox miRNAs have identifiable targets with near perfect complementarity (Molnar et al., 2007; Li et al., 2014; Voshall et al., 2014) and recent evidence suggests that miRNA regulation of transcript expression in Chlamydomonas may operate, at least for certain targets, by translation repression (Ma et al., 2013; Yamasaki et al., 2013; Voshall et al., 2014).

In summary, the RNA interference mechanism appears to be entirely absent from some microalgae and its biological role(s) in the species that possess core RNAi machinery components is poorly understood. Limited evidence suggests that a siRNA pathway may operate as a defense mechanism against transposon mobilization and, possibly, in antiviral immunity. A miRNA pathway, when present, may contribute to endogenous gene regulation. However, the identification of genuine miRNA targets remains challenging and, to date, no specific metabolic or physiological process controlled or modulated by miRNAs has been clearly defined in microalgae. Other possible functions of RNAi in phenomena such as heterochromatin formation, DNA methylation, DNA repair, or maintenance of genome stability, to our knowledge, have not been explored in these aquatic organisms.

5. Gene silencing mechanisms and biofuel/biomaterial production

Algae exhibit the potential for manufacturing a wide range of biofuel precursors and bioproducts, and advances in algal genomics, genetic engineering, and synthetic biology may facilitate the development of industrial strains suitable to specific production purposes. However, a serious limitation to strain improvement is our incomplete understanding of gene and metabolic network regulation in most algal species. In many eukaryotes, gene-silencing mechanisms have been implicated in context-dependent reversible gene repression, a critical component of gene networks involved in stress responses and in developmental programs (Law and Jacobsen, 2010; Bannister and Kouzarides, 2011; Ohsawa et al., 2013; Derkacheva and Hennig, 2014). Interestingly, most microalgae accumulate biofuel-

precursor compounds, such as starch and triacylglycerols (TAGs), primarily under stress conditions (Hu et al., 2008; Radakovits et al., 2010; La Russa et al., 2012; Liu and Benning, 2013). Transcriptome analyses in *C. reinhardtii* (reviewed by Liu and Benning, 2013) have revealed that a subset of genes involved in TAG biosynthesis, such as those encoding certain acyl-CoA: diacylglycerol acyltransferases, display greatly enhanced or almost exclusive expression in nitrogen-deprived cells. These observations suggest that gene-silencing mechanisms may indeed be involved in modulating metabolic responses to stress in microalgae. Specific genes for storage compound biosynthesis appear to be repressed under normal environmental conditions, when they may not be needed, and activated primarily under a variety of stresses. Yet, to our knowledge, direct evidence for a role(s) of epigenetic silencing mechanisms in microalgal gene regulation is currently missing.

Notwithstanding its endogenous role, RNA interference also provides a tool for functional genomic analyses in microalgae (reviewed by Cerutti et al., 2011). In some species, transient gene repression has been achieved by introduction into algal cells of non-integrative dsRNA/sRNA producing plasmids or of exogenously synthesized dsRNAs/sRNAs (Cerutti et al., 2011; Gimpel et al., 2013). In microalgae with available genome transformation methodologies stable and heritable RNAi has been developed, relying generally on the production of hairpin dsRNAs or artificial microRNAs from genome-integrated transgenes (Radakovits et al., 2010; Cerutti et al., 2011; Gimpel et al., 2013; Liu and Benning, 2013). These approaches have been used to suppress endogenous gene expression in microalgae, in order to characterize functionally genes involved in lipid metabolism (Gimpel et al., 2013; Liu and Benning, 2013) as well as to engineer algal strains with altered traits for biotechnological purposes, such as optimizing antenna size to improve light utilization and photosynthetic efficiency (Gimpel et al., 2013). RNAi technology (Cerutti et al., 2011), in conjunction with system level 'omics' approaches, may contribute much-needed insight into gene function, metabolic pathways, and regulatory networks in microalgae, as well as provide a valuable tool for the genetic engineering of enhanced traits.

6. Conclusion and perspective

Microalgae exhibit the potential for manufacturing biofuel precursors and a range of valuable bioproducts. However, a serious limitation to rational strain improvement is our incomplete understanding of gene function and regulation in most species. Gene-silencing mechanisms appear to be widespread in the Archaeplastida microalgae, including post-translational histone modifications, DNA methylation, and small RNA directed pathways. Yet, there also appears to be great diversity in the silencing systems present in individual algal species. Future challenges will involve defining whether these epigenetic processes regulate storage compound metabolic networks and developing improved tools for the genetic engineering of production strain.

Acknowledgments – This work was supported in part by a grant from the National Science Foundation (to H.C.). We also acknowledge the support of the Nebraska EPSCoR program and the Korea Research Institute of Bioscience and Biotechnology.

References

Allis, C.D., Berger, S.L., Cote, J., Dent, S., Jenuwien, T., Kouzarides, T., Pillus, L., Reinberg, D., Shi, Y., Shiekhattar, R., Shilatifard, A., Workman, J., Zhang, Y., 2007. New nomenclature for chromatin-modifying enzymes. Cell 131, 633–636.

Arabidopsis Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408, 796–815.

Axtell, M.J., 2013. Classification and comparison of small RNAs from plants. Annu. Rev. Plant Biol. 64, 137–159.

Babinger, P., Völkl, R., Cakstina, I., Maftei, A., Schmitt, R., 2007. Maintenance DNA methyltransferase (Met1) and silencing of CpG-methylated foreign DNA in Volvox carteri. Plant Mol. Biol. 63, 325–336.

Bannister, A.J., Kouzarides, T., 2011. Regulation of chromatin by histone modifications. Cell Res. 21, 381–395.

Bartel, D.P., 2009. MicroRNAs: target recognition and regulatory functions. Cell 136, 215–233.

Bhattacharya, D., Price, D.C., Chan, C.X., Qiu, H., Rose, N., Ball, S., Weber, A.P., Arias, M.C., Henrissat, B., Coutinho, P.M., Krishnan, A., Zäuner, S., Morath, S., Hilliou, F., Egizi, A., Perrineau, M.M., Yoon, H.S., 2013. Genome of the red alga *Porphyridium purpureum*. Nat. Commun. 4, 1941.

Blanc, G., Duncan, G., Agarkova, I., Borodovsky, M., Gurnon, J., Kuo, A., Lindquist, E., Lucas, S., Pangilinan, J., Polle, J., Salamov, A., Terry, A., Yamada, T., Dunigan, D.D., Grigoriev, Claverie, J.M., Van Etten, J.L., 2010. The *Chlorella variabilis* NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. Plant Cell 22, 2943–2955.

Blanc, G., Agarkova, I., Grimwood, J., Kuo, A., Brueggeman, A., Dunigan, D.D., Gurnon, J., Ladunga, I., Lindquist, E., Lucas, S., Pangilinan, J., Pröschold, T., Salamov, A., Schmutz, J., Weeks, D., Yamada, T., Lomsadze, A., Borodovsky, M., Claverie, J.M., Grigoriev, I.V., Van Etten, J.L., 2012. The genome of the polar eukaryotic microalga *Coccomyxa subellipsoidea* reveals traits of cold adaptation. Genome Biol. 13 (5), R39.

Burroughs, A.M., Ando, Y., Aravind, L., 2014. New perspectives on the diversification of the RNA interference system: insights from comparative genomics and small RNA sequencing. Wiley Interdiscip. Rev. RNA 5, 141–181.

Capuano, F., Mulleder, M., Kok, R., Blom, H.J., Ralser, M., 2014. Cytosine DNA methylation is found in *Drosophila melanogaster* but absent in *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and other yeast species. Anal. Chem. 86, 3697–3702.

Carthew, R.W., Sontheimer, E.J., 2009. Origins and mechanisms of miRNAs and siRNAs. Cell 136, 642–655.

Casas-Mollano, J.A., van Dijk, K., Eisenhart, J., Cerutti, H., 2007. SET3p monomethylates histone H3 on lysine 9 and is required for the silencing of tandemly repeated transgenes in *Chlamydomonas*. Nucleic Acids Res. 35, 939–950.

Casas-Mollano, J.A., Rohr, J., Kim, E.J., Balassa, E., van Dijk, K., Cerutti, H., 2008. Diversification of the core RNA interference machinery in *Chlamydomonas reinhardtii* and the role of DCL1 in transposon silencing. Genetics 179, 69–81.

Cerutti, H., Johnson, A.M., Gillham, N.W., Boynton, J.E., 1997. Epigenetic silencing of a foreign gene in nuclear transformants of *Chlamydomonas*. Plant Cell 9, 925–945.

Cerutti, H., Casas-Mollano, J.A., 2006. On the origin and functions of RNA-mediated silencing: from protists to man. Curr. Genet. 50, 81–99.

Cerutti, H., Ma, X., Msanne, J., Repas, T., 2011. RNA-mediated silencing in algae: biological roles and tools for the analysis of gene function. Eukaryot. Cell 10, 1164–1172.

Chisti, Y., 2013. Constraints to commercialization of algal fuels. J. Biotechnol. 167, 201–214.

Derelle, E., Ferraz, C., Rombauts, S., Rouzé, P., Worden, A.Z., Robbens, S., Partensky, F., Degroeve, S., Echeynié, S., Cooke, R., Saeys, Y., Wuyts, J., Jabbari, K., Bowler, C., Panaud, O., Piégu, B., Ball, S.G., Ral, J.P., Bouget, F.Y., Piganeau, G., De Baets, B., Picard, A., Delseny, M., Demaille, J., Van de Peer, Y., Moreau, H., 2006. Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. Proc. Natl. Acad. Sci. U.S.A. 103, 11647–11652.

Derkacheva, M., Hennig, L., 2014. Variations on a theme: polycomb group proteins in plants. J. Exp. Bot. 65, 2769–2784.

Du, J., Zhong, X., Bernatavichute, Y.V., Stroud, H., Feng, S., Caro, E., Vashisht, A.A., Terragni, J., Chin, H.G., Tu, A., Hetzel, J., Wohlschlegel, J.A., Pradhan, S., Patel, D.J., Jacobsen, S.E., 2012. Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell 151, 167–180.

Feng, S., Cokus, S.J., Zhang, X., Chen, P.Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., Halpern, M.E., Ukomadu, C., Sadler, K.C., Pradhan, S., Pellegrini, M., Jacobsen, S.E., 2010. Conservation and divergence of methylation patterning in plants and animals. Proc. Natl. Acad. Sci. U.S.A. 107, 8689–8694.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., Mello, C.C., 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature 391, 806–811.

Fortier, E., Belote, J.M., 2000. Temperature-dependent gene silencing by an expressed inverted repeat in *Drosophila*. Genesis 26, 240–244.

Ghildiyal, M., Zamore, P.D., 2009. Small silencing RNAs: an expanding universe. Nat. Rev. Genet. 10, 94–108.

Gimpel, J.A., Specht, E.A., Georgianna, D.R., Mayfield, S.P., 2013. Advances in microalgae engineering and synthetic biology applications for biofuel production. Curr. Opin. Chem. Biol. 17, 489–495.

Goll, M.G., Bestor, T.H., 2005. Eukaryotic cytosine methyltransferases. Annu. Rev. Biochem. 74, 481–514.

Griffiths, M.J., Harrison, S.T.L., 2009. Lipid productivity as a key characteristic for choosing algal species for biodiesel production. J. Appl. Phycol. 21, 493–507.

Hu, Q., Sommerfeld, M., Jarvis, E., Ghirardi, M., Posewitz, M., Seibert, M., Darzins, A., 2008. Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. Plant J. 54, 621–639.

Huang, Y., Liu, C., Shen, W.-H., Ruan, Y., 2011. Phylogenetic analysis and classification of the *Brassica rapa* SET-domain protein family. BMC Plant Biol. 11, 175.

Huff, J.T., Zilberman, D., 2014. Dnmt1-independent CG methylation contributes to nucleosome positioning in diverse eukaryotes. Cell 156, 1286–1297.

Krauss, V., 2008. Glimpses of evolution: heterochromatic histone H3K9 methyltransferase left its mark behind. Genetica 133, 93–106.

Krishnan, S., Horowitz, S., Trievel, R.C., 2011. Structure and function of histone H3 lysine 9 methyltransferases and demethylases. ChemBioChem 12, 254–263.

La Russa, M., Bogen, C., Uhmeyer, A., Doebbe, A., Filippone, E., Kruse, O., Mussgnug, J.H., 2012. Functional analysis of three type-2 DGAT homologue genes for triacylglycerol production in the green microalga *Chlamydomonas reinhardtii*. J. Biotechnol. 162, 13–20.

Law, J.A., Jacobsen, S.E., 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. 11, 204–220.

Lee, D.H., 2011. Algal biodiesel economy and competition among biofuels. Bioresour. Technol. 102, 43–49.

Leite, G.B., Abdelaziz, A.E., Hallenbeck, P.C., 2013. Algal biofuels: challenges and opportunities. Bioresour. Technol. 145, 134–141.

Li, J., Wu, Y., Qi, Y., 2014. MicroRNAs in a multicellular green alga *Volvox carteri*. Sci. China Life Sci. 57, 36–45.

Liu, B., Benning, C., 2013. Lipid metabolism in microalgae distinguishes itself. Curr. Opin. Biotechnol. 24, 300–309.

Ma, X., Kim, E.J., Kook, I., Ma, F., Voshall, A., Moriyama, E., Cerutti, H., 2013. Small interfering RNA-mediated translation repression alters ribosome sensitivity to inhibition by cycloheximide in *Chlamydomonas reinhardtii*. Plant Cell 25, 985–998.

Matsuzaki, M., Misumi, O., Shin-I, T., Maruyama, S., Takahara, M., Miyagishima, S.Y., Mori, T., Nishida, K., Yagisawa, F., Nishida, K., Yoshida, Y., Nishimura, Y., Nakao, S., Kobayashi, T., Momoyama, Y., Higashiyama, T., Minoda, A., Sano, M., Nomoto, H., Oishi, K., Hayashi, H., Ohta, F., Nishizaka, S., Haga, S., Miura, S., Morishita, T., Kabeya, Y., Terasawa, K., Suzuki, Y., Ishii, Y., Asakawa, S., Takano, H., Ohta, N., Kuroiwa, H., Tanaka, K., Shimizu, N., Sugano, S., Sato, N., Nozaki, H., Ogasawara, N., Kohara, Y., Kuroiwa, T., 2004. Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. Nature 428, 653–657.

Meister, G., 2013. Argonaute proteins: functional insights and emerging roles. Nat. Rev. Genet. 14, 447–459.

Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B., Terry, A., Salamov, A., Fritz-Laylin, L.K., Maréchal-Drouard, L., Marshall, W.F., Qu, L.H., Nelson, D.R., Sanderfoot, A.A., Spalding, M.H., Kapitonov, V.V., Ren, Q., Ferris, P., Lindquist, E., Shapiro, H., Lucas, S.M., Grimwood, J., Schmutz, J., Cardol, P., Cerutti, H., Chanfreau, G., Chen, C.L., Cognat, V., Croft, M.T., Dent, R., Dutcher, S., Fernández, E., Fukuzawa, H., González-Ballester, D., González-Halphen, D., Hallmann, A., Hanikenne, M., Hippler, M., Inwood, W., Jabbari, K., Kalanon, M., Kuras, R., Lefebvre, P.A., Lemaire, S.D., Lobanov, A.V., Lohr, M., Manuell, A., Meier, I., Mets, L., Mittag, M., Mittelmeier, T., Moroney, J.V., Moseley, J., Napoli, C., Nedelcu, A.M., Niyogi, K., Novoselov, S.V., Paulsen, I.T., Pazour, G., Purton, S., Ral, J.P., Riaño-Pachón, D.M., Riekhof, W., Rymarquis, L., Schroda, M., Stern, D., Umen, J., Willows, R., Wilson, N., Zimmer, S.L., Allmer, J., Balk, J., Bisova, K., Chen, C.J., Elias, M., Gendler, K., Hauser, C., Lamb, M.R., Ledford, H., Long, J.C., Minagawa, J., Page, M.D., Pan, J., Pootakham, W., Roje, S., Rose, A., Stahlberg, E., Terauchi, A.M., Yang, P., Ball, S., Bowler, C., Dieckmann, C.L., Gladyshev, V.N., Green, P., Jorgensen, R., Mayfield, S., Mueller-Roeber, B., Rajamani, S., Sayre, R.T., Brokstein, P., Dubchak, I., Goodstein, D., Hornick, L., Huang, Y.W., Jhaveri, J., Luo, Y., Martínez, D., Ngau, W.C., Otillar, B., Poliakov, A., Porter, A., Szajkowski, L., Werner, G., Zhou, K., Grigoriev, I.V., Rokhsar, D.S., Grossman, A.R., 2007. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. Science 318, 245–251.

Molnar, A., Schwach, F., Studholme, D.J., Thuenemann, E.C., Baulcombe, D.C., 2007. MiRNAs control gene expression in the single-cell alga *Chlamydomonas reinhardtii*. Nature 447, 1126–1129.

Moreau, H., Verhelst, B., Couloux, A., Derelle, E., Rombauts, S., Grimsley, N., Van Bel, M., Poulain, J., Katinka, M., Hohmann-Marriott, M.F., Piganeau, G., Rouzé, P., Da Silva, C., Wincker, P., Van de Peer, Y., Vandepoele, K., 2012. Gene functionalities and genome structure in *Bathycoccus prasinos* reflect cellular specializations at the base of the green lineage. Genome Biol. 13, R74.

Mujtaba, S., Manzur, K.L., Gurnon, J.R., Kang, M., Van Etten, J.L., Zhou, M.M., 2008. Epigenetic transcriptional repression of cellular genes by a viral SET protein. Nat. Cell Biol. 10, 1114–1122.

Nishiyama, R., Wada, Y., Mibu, M., Yamaguchi, Y., Shimogawara, K., Sano, H., 2004. Role of a non-selective de novo DNA methyltransferase in maternal inheritance of chloroplast genes in the green alga *Chlamydomonas reinhardtii*. Genetics 168, 809–816.

Ohsawa, R., Seol, J.H., Tyler, J.K., 2013. At the intersection of non-coding transcription, DNA repair, chromatin structure, and cellular senescence. Front. Genet. 4, 136.

Oliver, C., Santos, J.L., Pradillo, M., 2014. On the role of some ARGONAUTE proteins in meiosis and DNA repair in *Arabidopsis thaliana*. Front. Plant Sci. 5, 177.

Palenik, B., Grimwood, J., Aerts, A., Rouzé, P., Salamov, A., Putnam, N., Dupont, C., Jorgensen, R., Derelle, E., Rombauts, S., Zhou, K., Otillar, R., Merchant, S.S., Podell, S., Gaasterland, T., Napoli, C., Gendler, K., Manuell, A., Tai, V., Vallon, O., Piganeau, G., Jancek, S., Heijde, M., Jabbari, K., Bowler, C., Lohr, M., Robbens, S., Werner, G., Dubchak, I., Pazour, G.J., Ren, Q., Paulsen, I., Delwiche, C., Schmutz, J., Rokhsar, D., Van de Peer, Y., Moreau, H., Grigoriev, I.V., 2007. The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. Proc. Natl. Acad. Sci. U.S.A. 104, 7705–7710.

Patrick, K.L., Shi, H., Kolev, N.G., Ersfeld, K., Tschudi, C., Ullu, E., 2009. Distinct and overlapping roles of two Dicer-like proteins in the RNA interference pathways of the ancient eukaryote *Trypanosoma brucei*. Proc. Natl. Acad. Sci. U.S.A. 106, 17933–17938.

Ponger, L., Li, W.-H., 2005. Evolutionary diversification of DNA methyltransferases in eukaryotic genomes. Mol. Biol. Evol. 22, 1119–1128.

Price, D.C., Chan, C.X., Yoon, H.S., Yang, E.C., Qiu, H., Weber, A.P., Schwacke, R., Gross, J., Blouin, N.A., Lane, C., Reyes-Prieto, A., Durnford, D.G., Neilson, J.A., Lang, B.F., Burger, G., Steiner, J.M., Löffelhardt, W., Meuser, J.E., Posewitz, M.C., Ball, S., Arias, M.C., Henrissat, B., Coutinho, P.M., Rensing, S.A., Symeonidi, A., Doddapaneni, H., Green, B.R., Rajah, V.D., Boore, J., Bhattacharya, D., 2012. *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. Science 335, 843–847.

Prochnik, S.E., Umen, J., Nedelcu, A.M., Hallmann, A., Miller, S.M., Nishii, I., Ferris, P., Kuo, A., Mitros, T., Fritz-Laylin, L.K., Hellsten, U., Chapman, J., Simakov, O., Rensing, S.A., Terry, A., Pangilinan, J., Kapitonov, V., Jurka, J., Salamov, A., Shapiro, H., Schmutz, J., Grimwood, J., Lindquist, E., Lucas, S., Grigoriev, I.V., Schmitt, R., Kirk, D., Rokhsar, D.S., 2010. Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. Science 329, 223–226.

Radakovits, R., Jinkerson, R.E., Darzins, A., Posewitz, M.C., 2010. Genetic engineering of algae for biofuel production. Eukaryot. Cell 9, 486–501.

Rajakumara, E., Law, J.A., Simanshu, D.K., Voigt, P., Johnson, L.M., Reinberg, D., Patel, D.J., Jacobsen, S.E., 2011. A dual flip-out mechanism for 5mC recognition by the Arabidopsis SUVH5 SRA domain and its impact on DNA methylation and H3K9 dimethylation in vivo. Genes Dev. 25, 137–152.

Rogers, K., Chen, X., 2013. Biogenesis, turnover, and mode of action of plant miRNAs. Plant Cell 25, 2383–2399.

Rowe, J.M., Dunigan, D.D., Blanc, G., Gurnon, J.R., Xia, Y., Van Etten, J.L., 2013. Evaluation of higher plant virus resistance genes in the green alga, *Chlorella variabilis* NC64A, during the early phase of infection with *Paramecium bursaria* chlorella virus-1. Virology 442, 101–113.

Saze, H., Kakutani, T., 2011. Differentiation of epigenetic modifications between transposons and genes. Curr. Opin. Plant Biol. 14, 81–87.

Schönknecht, G., Chen, W.H., Ternes, C.M., Barbier, G.G., Shrestha, R.P., Stanke, M., Bräutigam, A., Baker, B.J., Banfield, J.F., Garavito, R.M., Carr, K., Wilkerson, C., Rensing, S.A., Gagneul, D., Dickenson, N.E., Oesterhelt, C., Lercher, M.J., Weber, A.P., 2013. Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. Science 339, 1207–1210.

Shaver, S., Casas-Mollano, J.A., Cerny, R.L., Cerutti, H., 2010. Origin of the polycomb repressive complex 2 and gene silencing by an E(z) homolog in the unicellular alga *Chlamydomonas*. Epigenetics 5, 301–312.

Strenkert, D., Schmollinger, S., Schroda, M., 2013. Heat shock factor 1 counteracts epigenetic silencing of nuclear transgenes in *Chlamydomonas reinhardtii*. Nucleic Acid Res. 41, 5273–5289.

Szittya, G., Silhavy, D., Molnár, A., Havelda, Z., Lovas, A., Lakatos, L., Bánfalvi, Z., Burgyán, J., 2003. Low temperature inhibits RNA silencing-mediated defence by the control of siRNA generation. EMBO J. 22, 633–640.

Thomas, M., Lieberman, J., Lai, A., 2010. Desperately seeking microRNA targets. Nat. Struct. Mol. Biol. 17, 1169–1174.

Tirichine, L., Bowler, C., 2011. Decoding algal genomes: tracing back the history of photosynthetic life on Earth. Plant J. 66, 45–57.

Umen, J.G., Goodenough, U.W., 2001. Chloroplast DNA methylation and inheritance in *Chlamydomonas*. Genes Dev. 15, 2585–2597.

Voshall, A., Kim, E.J., Ma, X., Moriyama, E.N., Cerutti, H., 2014. Identification of AGO3 associated miRNAs and computational prediction of their targets in the green alga *Chlamydomonas reinhardtii*. Genetics (submitted for publication).

Wei, H., Zhou, M.M., 2010. Viral-encoded enzymes that target host chromatin functions. Biochim. Biophys. Acta 1799, 296–301.

Worden, A.Z., Lee, J.H., Mock, T., Rouzé, P., Simmons, M.P., Aerts, A.L., Allen, A.E., Cuvelier, M.L., Derelle, E., Everett, M.V., Foulon, E., Grimwood, J., Gundlach, H., Henrissat, B., Napoli, C., McDonald, S.M., Parker, M.S., Rombauts, S., Salamov, A., Von Dassow, P., Badger, J.H., Coutinho, P.M., Demir, E., Dubchak, I., Gentemann, C., Eikrem, W., Gready, J.E., John, U., Lanier, W., Lindquist, E.A., Lucas, S., Mayer, K.F., Moreau, H., Not, F., Otillar, R., Panaud, O., Pangilinan, J., Paulsen, I., Piegu, B., Poliakov, A., Robbens, S., Schmutz, J., Toulza, E., Wyss, T., Zelensky, A., Zhou, K., Armbrust, E.V., Bhattacharya, D., Goodenough, U.W., Van de Peer, Y., Grigoriev, I.V., 2009. Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes micromonas. Science 324, 268–272.

Worden, A.Z., Allen, A.E., 2010. The voyage of the microbial eukaryote. Curr. Opin. Microbiol. 13, 652–660.

Yamasaki, T., Voshall, A., Kim, E.J., Moriyama, E., Cerutti, H., Ohama, T., 2013. Complementarity to an miRNA seed region is sufficient to induce moderate repression of a target transcript in the unicellular green alga *Chlamydomonas reinhardtii*. Plant J. 76, 1045–1056.

Zemach, A., McDaniel, I.E., Silva, P., Zilberman, D., 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 328, 916–919.

Zhao, T., Li, G., Mi, S., Li, S., Hannon, G.J., Wang, X.J., Qi, Y., 2007. A complex system of small RNAs in the unicellular green alga *Chlamydomonas reinhardtii*. Genes Dev. 21, 1190–1203.

Zheng, S., Wang, J., Feng, Y., Wang, J., Ye, K., 2012. Solution structure of MSL2 CXC domain reveals an unusual Zn3Cys9 cluster and similarity to Pre-SET domains of histone lysine methyltransferases. PLoS One 7, e45437.

Zhong, X., Du, J., Hale, C.J., Gallego-Bartolome, J., Feng, S., Vashisht, A.A., Chory, J., Wohlschlegel, J.A., Patel, D.J., Jacobsen, S.E., 2014. Molecular mechanism of action of plant DRM de novo DNA methyltransferases. Cell 157, 1050–1060.